ABSTRACT

Web sites are analyzed to determine whether the web sites are embedding session identifiers in web documents. The analysis is based on a comparison of in-host links of multiple copies of a document from a web site. Rules governing the insertion of session identifiers for the web site may be determined and used to assist in crawling the web site.